DSSTox Field Definition File:

EPA Water Disinfection By-Products with Carcinogenicity Estimates (DBPCAN)

(last updated 15 February 2008)

Description: Information in this file is intended to provide a minimum level of annotation to the DSSTox SDF (Structure Data Format) file created for the Water Disinfection By-Products with Carcinogenicity Estimates database (DBPCAN) abstracted from the Main Citation listed below (Woo et al., 2002), and supplemented by additional information provided by the main authors of that study. For further explanation of Source-specific fields and background pertaining to the content of this database, a user is encouraged to consult the Main Citation below. Additional information is provided on the DSSTox DBPCAN SDF Download Page http://www.epa.gov/nheerl/dsstox/sdf_dbpcan.html. This database differs significantly from other DSSTox databases in that the chemical substances did not undergo actual toxicity testing; rather each of the included molecular structures was elucidated by analytical and spectroscopic techniques as being present in drinking water samples treated by various disinfection methods, and for each of the included structures a prediction of carcinogenic potential was generated. Hence, the term "TestSubstance" refers here to an SAR prediction rather than an experimental determination. Another important clarification in relation to this database is to distinguish two categories of rationale provided in support of the carcinogenic potential predictions in the **ActivityConcernLevel_Rationale** field. The first category consists of the more extensively documented and referenced rationales provided in the body of the Main Citation (Woo et al., 2002). For these chemicals, the **ActivityConcernLevel_RationaleSource** field entry refers to the corresponding Table in the Main Citation. For the remaining structures in the database, less well documented rationale was provided by the authors for DSSTox use; this category of unpublished rationale is denoted "author communication".

Description of ***DSSTox Standard Chemical Fields*** can be found in the Central Field Definition Table located at:
http://www.epa.gov/ncct/dsstox/CentralFieldDef.html

The first section of the table below lists the ***DSSTox Standard Toxicity Fields*** employed for this database, followed by the ***DBPCAN Source-Specific Fields*** containing the toxicity information particular to DBPCAN. The ***Field Type*** indicates the type of the field, such as numeric, integer, defined text, memo, etc. All ***Units*** and ***Descriptions*** are extracted from Source reference materials unless otherwise noted. ***Allowable Entries*** lists allowed field entries occurring in DBPCAN, separated by slashes for exclusive entries (i.e., cannot occur with another entry) and semicolons or spaces for non-exclusive entries (i.e., can occur with other values). These are defined and explained in the ***Description*** section.

Source Contact: Scientific questions pertaining to this database should be directed to Yin-tak Woo, Risk Assessment Division, Office of Pollution Prevention & Toxics, US EPA, Wash. DC; email: woo.yintak@epa.gov

Main Citation: Publications reporting use of the DBPCAN DSSTox SDF file are asked to list the full DSSTox file name(s), including date stamp, and to cite as primary reference the following:

Woo, Y.T., D. Lai, J.L. McLain, M.K. Manibusan, and V. Dellarco (2002) Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products, *Environ Health Perspect,*110 Suppl 1: 75-87. *

**pdf of Main Citation can be downloaded from DSSTox DBPCAN Database page:* http://www.epa.gov/ncct/dsstox/sdf_dbpcan.html

SDF Usage Notes:
Each DSSTox SDF file contains a single **STRUCTURE** field. For each chemical record, the **STRUCTURE** field entry directly corresponds to the content of the **STRUCTURE_...** fields. The **STRUCTURE_Shown** field documents the relationship between what is displayed in the **STRUCTURE** field and the actual tested chemical substance, i.e. **TestSubstance_...** fields, with the latter corresponding directly to the toxicity data field entries. Commercial chemical relational database (CRD) applications may automatically insert one or more structure identifier fields upon import or export of an SDF file (e.g., Formula, FW or Mol_ID), fields that may augment or duplicate one or more of the DSSTox Standard Chemical Fields. Users are cautioned that fields containing null values in the first record of the SDF will be reordered upon import into most applications; for this reason, the word "blank" has been inserted into null fields in Record 1 of DSSTox SDF files and can be deleted after SDF import. Users are additionally cautioned that some fields (**STRUCTURE_SMILES** and **STRUCTURE_InChI**, in particular) may exceed the 200 character limit specified in the MDL CTFiles SDF standard (see http://www.epa.gov/ncct/dsstox/MoreonSDF.html), and that some CRD applications may insert a line break or truncate these fields upon SDF import or export. Finally, CRD application-specific molecular header information in the SDF file is deleted in the final DSSTox SDF files; users running CRD applications requiring a unique molecule header upon import of the SDF can specify either **DSSTox_RID** or the **DSSTox_FileID** be used. Upon SDF import, **DSSTox_CID** can be used to identify and manage chemical structure duplicates and **DSSTox_Generic_SID** can be used to identify common Test Substances across and within DSSTox files (similar to CASRN-substance, but available for all DSSTox substances and further distinguishes among different purity/grade substances).

As an MS Word document, the following table is best viewed onscreen using either Normal or Web Layout View in Landscape page orientation.

Field Name	***Field Type***	***Units***	***Allowable Entries***	***Description***	***Comments***
			DSSTox Standard Toxicity Fields		
Study Type *(no spaces)*	defined text		SAR Prediction	Field is used to label all records in the database, generally with the same entry for all records, and is designed to facilitate record identification for cross-database structure searching. Field entry refers to the main type of toxicity study for which data is represented in the database.	Field names and content are being coordinated with the public ToxML standardization effort.
Endpoint	defined text		Carcinogenicity	Field is used to label all records in the database, generally with the same entry for all records, and is designed to facilitate record identification for cross-database structure searching. Field entry refers to the type of toxicity measure represented within the database.	Field names and content are being coordinated with the public ToxML standardization effort.
			DBPCAN Source-Specific Fields		
ChemClass_DBP	defined text		Acetate of haloalcohols/ Haloacids/ Haloaldehydes/ Haloamines and haloamides/ Haloethers/ Halofuranones and related compounds/ Halogenated aromatics/ Haloalkanes and haloalkenes/ Haloketones/ Halonitriles/ Halonitroalkanes/ Inorganics/ Nonhalogenated ketones/ Nonhalogenated aldehydes/ Nonhalogenated acids/ Nonhalogenated aromatics/	Chemical classification categories considered in analog searches and in developing structure-activity relationship (SAR) rationales for predicting carcinogenic potential rankings.	These entries correspond to "Chemical class" entries in Table 3 of Main Citation. Haloalkanes/alkenes converted to Haloalkanes and haloalkenes

			Other halogenated organics/ Other nonhalogenated organics/		

ActivityOutcome _DBPCAN *(no spaces)*	defined text		active/ inactive/ inconclusive/	Activity Outcome maps **ActivityConcernLevel_Carcinogenicity** as follows: "active" = High, High-Moderate, Moderate, or Low-Moderate; "inconclusive" = Marginal; "inactive" = Low.	Summary activity call for use in PubChem and structure-activity relationship studies. *New field added to v4b.*
ActivityScore _DBPCAN *(no spaces)*	integer		INTEGER [0-90]	Activity Score is assigned based on **ActivityConcernLevel _Carcinogenicity** as follows (using highest route of exposure estimate): High=90; High-Moderate=70; Moderate=50; Low-Moderate=30; Marginal=10; Low=0.	Summary activity call for use in PubChem and structure-activity relationship studies. *New field added to v4b.*
ActivityConcern Level _Carcinogenicity *(no spaces)*	defined text		High/ High-Moderate/ Moderate/ Low-Moderate/ Marginal/ Low/	Concern level predictions are based on expert judgment relative to known carcinogens and using principles of mechanism-based structure-activity analysis. Factors taken into consideration include structural analogy to known carcinogens, toxicokinetic and toxicodynamic factors, potency indicators for a structural analog (such as multispecies, multitarget or multisite carcinogens), short-term test data, and metabolic activation. Concern levels are: High = highly likely to be a potent multispecies, multitarget carcinogen even at low doses; High-Moderate = highly likely to be an active multispecies/target carcinogen even at moderate doses; Moderate = likely to be a moderately active multispecies/target carcinogen at relatively high doses or active single species/target carcinogen at low doses; Low-Moderate = likely to be weakly carcinogenic, or carcinogenic toward a single species/target at relatively high doses; Marginal = likely to have marginal carcinogenic activity or may be weakly carciongenic at doses at or exceeding maximum tolerated doses; Low = unlikely to be carcinogenic.	Corresponds to entries in Table 3 of Main Citation.
ActivityConcern Level_Rationale *(no spaces)*	memo		*Text*	Concise narrative statement summarizing evidence supporting the prediction of carcinogenic potential for the DBP chemical. Rationale is derived from mechanism-based SAR analysis, and is strongly reliant on identification of one or a few close structural analogs with known carcinogenicity and supplemented by extensive literature search for genotoxicity and other data.	Two categories of rationale are provided as indicated by **ActivityConcernLevel_RationaleSource** entry. Rationale narratives exceeding the field limit of 255 characters are abbreviated and identified in the **Note_DBPCAN** field.
ActivityConcern Level _RationaleSource *(no spaces)*	defined text		Table 5/ Table 6/ Table 7/ Table 9/ author communication/	Source of rationale narrative, either from Main Citation (Tables 5,6,7, or 9) or from supplemental material provided by author communication.	Two categories of rationale are provided: 1) rationales from Tables 5,6,7, or 9 of Main Citation include references and greater documentation; 2) rationales provided by "author communication" have no references and

					less documentation.
Analog_Chemical Name *(no spaces)*	memo		*Text*	Chemical name of primary structural analog cited in **ActivityConcernLevel_Rationale** for SAR carcinogenic potential prediction listed in Table 1.	
Analog_CASRN	text		NOCAS/ ######-##-#	CASRN of primary structural analog cited in SAR rationale for carcinogenic potential prediction, corresponding to **Analog_ChemicalName**.	
Analog_SMILES	memo		*Text*	SMILES code of primary structural analog cited in SAR rationale for carcinogenic potential prediction, corresponding to **Analog_ChemicalName**.	
Note_DBPCAN	memo		*Text*	Field used to provide supplementary Source-specific information pertaining to the chemical and toxicity fields.	DBPCAN-specific information pertaining to the chemical record previously included in **ChemicalNote** field has been moved to this field. Chemical name synonyms were formerly provided in retired field: **TestSubstance_ChemicalName_Other**. Chemical name synonyms provided here if offered in original Source materials National occurrence data collected for this DBP chemical under the US EPA information collection rule (see Table 1 and Ref. 8 in Main Citation). Rationale condensed = Narrative published in Main Citation was condensed in order not to exceed field limit of 255 characters. Replaces **ToxicityNote** field (May 2007).